International Conference on Applied Internet and Information Technologies, 2016
DOI:10.20544/AIIT2016.01

MongoDB as solution for data warehousing in online HRM systems

Biljana Radulović[1], Dragica Radosav[1], and Milan Malić[1]

[1] University of Novi Sad / Technical Faculty "Mihajlo Pupin", Djure Djakovica bb, 23000 Zrenjanin, Serbia
biljana.radulovic66@gmail.com; radosav@tfzr.uns.ac.rs; milanmalic@outlook.com

Abstract. With development of technology and techniques for data collection, modern business organizations are able to store large amounts of new information about customers and employees within it. In order to respond to the needs identified in the field of data warehousing and data analysis within human enterprise resource planning (ERP), business organizations increasingly opt for new ways to store the data. One of them is use of non-relational databases (NoSQL). This paper presents an information system (IS), developed for the needs of human resource management (HRM). For data warehousing IS use benefits of NoSQL MongoDB and was developed to be accessible via the Internet At this point, the authors aim to point out the potential solution in data warehousing and ERP system development. This solution can be used as the initial point for the systems further evolving, which tend to support futures that Industry 4.0 offers.

Keywords: NoSQL, HRM, IS development, data analysis.

1. Introduction

The emergence of Web 2.0 and its influence on the development of modern web applications, have brought about a major shift in the way of running the business organizations [1]. Nowadays, the Internet as we know it and the data exchanged on it, significantly differ from their original purpose. The amount of data and their volume exchanged over the Internet are nowadays measured in zettabytes per year, with the increase of 2 zettabytes yearly, until 2019 [2].

According to the research conducted in 2009, only a small number of worldwide business organizations used the Big Data while running their projects, and the total amount of material resources invested in the projects did not exceed 100 million dollars [3]. Today, a lot of companies invest heavily in development of Big Data. This fact has been pointed out by David White, the CEO of Import.io:

"Traditionally, Enterprise has heavily invested in Big Data teams and infrastructure, but what we are seeing now is democratization of Big Data – the tools/services to collect and analyse data at scale are now within the price range of mainstream business. It's the start of something very exciting".

A very important conclusion can be derived from the enclosed information, supporting the fact that the business organizations are ready to change and accept the

new, so far unknown ways of storing, analysing and processing data in order to reach the new findings related to their business. The development of new systems and application of new technologies in the domain of data warehousing can lead to significant benefits, from more efficient data management, faster and more precise analysis, to delivering the potentially critically important data in real time.

It is the ERP systems that have been researched and improved over years, using numerous new technologies, like innovations in the field of communication technologies, service-oriented architecture [4], etc. However, regardless of all the innovation mentioned and improvements applied, they still can be made better. Therefore, one could say that ERP systems can be improved today through the implementation of new data warehousing technologies, like non-relational databases (NoSQL).

NoSQL databases represent a database group which is not based on the relational model. The data model, used for warehousing by NoSQL database, is not fixed, but a frequent possibility of NoSQL database, which is not used by the relational database model. In the majority of cases, in fact, NoSQL databases do use SQL, so NoSQL does not mean that SQL is absolutely not used. Therefore, we can see from the literature that NoSQL also means „Not only SQL“[6].

By using all the mentioned technologies, ERP systems can be significantly improved, so as to be made accessible in real time to various groups of people within the organization, like management, economists, analysts, marketing professionals and such. In this way, the time needed for making decisions, as well as the number of communication and decision-making errors have been greatly reduced. Real-time data processing and displaying the ones that the management was unaware of, enhances the managing of an organization and its resources.

The rest of the paper is organized as follows: Section II. Review of the papers in the field, Section III. Detailed review of the suggested system functionality, Section IV. Development, technologies and description of the system architecture, Section V. Conclusions and future research.

2. Related work

ERP systems have a significant role in making business decisions. They improve business and organization efficiency. It can be said that there are numerous papers and applicable approaches in the field of ERP system developing. However, it should be pointed out that there are a really small number of the ones including NoSQL databases as the data warehousing mechanism in their development [5].

Recent research, conducted by some media companies and regarding the expansion of NoSQL, point out the annual growth of 21%, in the period from 2013 to 2018 [7]. Hence, it can be said that exactly this type of growth promotes the technology as one of the fastest growing.

The increasing application of NoSQL is reflected in the numerous benefits it brings [8-12]. NoSQL databases can be easily scaled, enabling a flexible data model [10][11]. Thanks to their horizontal scalability, NoSQL databases do not have limits, on the contrary to the vertical, which adds up new server clusters [12]. In this way, greater capacity is obtained and there is no performance loss. Thanks to their flexible scheme,

on the contrary to the relational databases, NoSQL databases do not require scheme defining nor flexible dynamic schemes [13]. The asynchronous replication is used when creating and storing the data copies, enabling faster creation of the documents stored in NoSQL database, and thus faster work.

Anyway, flaws of NoSQL should be pointed out as well. As the author of the paper [14] says, NoSQL enables BASE, on the opposite to SQL, enabling ACID. A more recent research indicates possible solutions to this problem, suggesting that they should be made on the middle tier in order to support the ACID features in NoSQL database [15].

Thanks to their increasing popularity, a large number of NoSQL database types have been developed in the recent years. They can be divided into several groups according to the way of storing. Each of these groups has a large number of various producers developing and maintaining their databases. They mainly differ mutually in a certain segment within their group, and each of them has its advantages and disadvantages [8][13].

The papers published in the several recent years show that the NoSQL databases have been developed specifically for the needs of Big Data [8][11][16]. According to the research in the field, it can be said with certainty that NoSQL will remain the field leader [7].

In the paper [17], the author points out the need to invest into innovative technologies, web services as well as the technology pervasion, since they will be of vital significance if the producers of ERP systems want to survive in the future. For this reason, it is necessary to accept new trends, which, as already mentioned, include the implementation of NoSQL.

One of the important aspects of modern ERP systems is definitely the data transfer between users and remote company divisions, in real time. The paper [18] points out the possible solutions regarding improvement of applications to enable real-time information delivery. It can be easily concluded that this is really significant since real-time data exchange lead to more efficient decision making when problems occur in a business organization's work, and application of new technologies definitely leads to more precise and clear answers.

3. Functionality

From the user's viewpoint, the solution for managing human resources consists of a web application, which can be accessed over the Internet.

This application is in fact an ERP system, designed so as to meet all the requirements of human resource management, in order to remotely collect all the relevant information on the employees' work. It creates the records of the employees' visits and activities throughout the workday, their performance and effective work time. In this way, the business organization can collect and analyse their employees' work, and then establish a more efficient way of their management. Based on the processed data, the organization management can make more effective decisions regarding allocation of human resources on active projects, and all with the aim to reduce costs and increase work efficiency.

Given that the application was developed for the web environment, using the program languages PHP and JavaScript, and its Graphical User Interface (GUI) in HTML5 format, it can be accessed from all the contemporary portable devices which support the most popular browsers.

The major benefit of this ERP system for human resource management, is the fact that it uses NoSQL MongoDB database for warehousing all the collected information. In this way, a possibility to create applications bringing the new data warehousing solution for business organizations is pointed out. As it has been presented in the paper, NoSQL brings multiple benefits that can tackle the majority of new problems, regarding storing the growing data.

Fig. 1. Main control panel. This figure shows the control panel is segmented into fore mentioned modules, and each of them has the option of data entry, data display and data removal from NoSQL database.

The application consists of three parts – control panel, user interface and front-end application. In this way, data managing is separated and safety is increased. The modules of the application's control panel are as follows:

- Simple configuration
- Creating the user groups and their allocation to employees
- Creating a detailed employees' activity log throughout the day
- Client management
- Employee management
- Creating, managing and monitoring of the project
- Company structure creating
- Workplace management
- Managing salary scales
- Travel management
- Salary management
- Loan management
- Creating reports

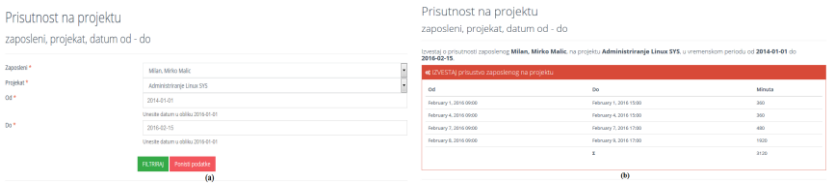

Fig. 2. Consists of two parts, where (a) shows the data entry form, on the basis of which the report will be created, and (b) shows the created report. As it can be seen in Figure 2a, it is possible to filter the reports by multiple parameters, including the date range. Figure 2b presents the use of complex aggregate queries on NoSQL databases, to obtain the result set.

As of the user interface, its first role is the automated collection of employees' activities during a session. The modules of the interface are as follows:

- Collecting the data on employees' project activities
- Displaying the information on employees' project activities
- Displaying the information on the previous period activities
- Displaying the information on travel and finances
- Payroll accounting for the current month, on the basis of the realized activities

Prisustvo zaposlenog za mesec February, 2016

Projekat	Prijava	Odjava	Napomena	Minuta
	February 11, 2016 05:38	February 11, 2016 05:40	Dodavanje putem korisnikovog panela	2.07
Administriranje Linux SYS	February 1, 2016 09:00	February 1, 2016 15:00	admin panel	360
E-shop computer store	February 2, 2016 09:00	February 2, 2016 15:00	priprema	360
E-shop computer store	February 3, 2016 09:00	February 3, 2016 15:00	priprema	360
Administriranje Linux SYS	February 4, 2016 09:00	February 4, 2016 15:00	priprema	360
E-shop computer store	February 5, 2016 09:00	February 5, 2016 15:00	priprema	360
Programiranje IS	February 6, 2016 09:00	February 6, 2016 15:00	analiza	360
Administriranje Linux SYS	February 7, 2016 09:00	February 7, 2016 17:00	administracija	480
Administriranje Linux SYS	February 8, 2016 09:00	February 9, 2016 17:00	administracija	1920
E-shop computer store	February 15, 2016 09:00	February 22, 2016 17:00	Razvoj FW	10560
			Σ minuta	15122.07

Obracun plate za mesec February, 2016

Sifra	Ostvreno radno vreme	Iznos	Valuta
Osnovica plate	15122.07	10000	evro
Sifra	**Opis**	**Iznos**	**Valuta**
Bonus I	Vise od 10200 minuta aktivnosti [bonus 5%]	500	evro
Bonus II	Vise od 2 aktivna projekta [bonus 2%]	0	evro
	Za isplatu Σ		evro

Fig. 3. Salary reports for employee. As it can be seen in Figure 3, apart from collecting information about the employees and their activities, the application for human resource management offers their overview as well, when creating the salary reports. In this way, the employees can have the insight in all their activities, and thus the ways to improve their work.

4. Development

The following techniques have been used for the developing of ERP system for human resource management, presented in Section III: NoSQL MongoDB [18], PHP program language, JavaScript, HTML5 and Linux Ubuntu Server 14.04 x64 for initiation of Apache2 web server.

MongoDB server was used for storing the data. It uses the object-structured objects, stored into collections as a data model [18]. As of queries, they are posed through manipulating the objects in the collections. A lot of people opt for using the MongoDB, as can be seen in the conducted research [19]. Easy use, flexible scheme and great scalability are just some of the benefits offered by this NoSQL database. Nowadays, a large number of leading organizations, like CERN, Craigslist, Foursquare etc. use MongoDB in their environment.

The document structure in MongoDB collection corresponds to JavaScript Object Notation (JSON) object. However, when storing the documents into the collection, MongoDB server performs the serialization of the object into the binary encoded system [18]. According to the research, BSON is a more efficient way to store data than XML and JSON format, due to the fact that it consumes fewer server resources, as well as the shorter time for processing. [20].

MongoDB-PHP Driver was used to connect MongoDB server and application. It enables API through which the connection with MongoDB server is established. At this point, it should be underlined that, in order for the driver to function properly, PHP 5.1 or up should be used. The driver is open-source and can be downloaded freely.

Regarding the document relation management, MongoDB as a non-relational model has no foreign key and JOIN operations. The connecting is done in the embedded documents or by creating a reference from one document to another [18].

Moreover, one of the great benefits of MongoDB lays in the fact that it can be used alongside with the RDBMS systems. It can be easily concluded from this fact, that it is a good solution for implementation with the already existing systems.

Additionally, MongoDB has the built-in system for managing big files, named GridFS. By using it, MongoDB stores the binary files into the database, but the fact that BSON object cannot be greater than 16MB should be pointed out.

5. Conclusions and future work

This paper presented the ways in which NoSQL DB can be successfully used to store data in running of ERP system. There are numerous benefits of using the NoSQL DB, from scalability, flexible scheme, work speed, smaller maintenance costs etc., which make these databases a new and competitive solution for developing modern web 2.0 applications.

Although a really small number of companies have been using Big Data in their projects only a couple of years ago, the things are rapidly changing nowadays and Fortune 500 companies are increasingly using them in some of their projects. In the same manner, the world-renown laboratories use NoSQL when conducting their experiments, to store the obtained results.

On the other hand, this technology also has some flaws. NoSQL does not provide ACID properties in their transactions, as it is the case with relational database management system. NoSQL provides BASE transaction properties, as presented in this paper. However, as mentioned, the research shows that these flaws can be overcome.

The increasing popularity of NoSQL DB in active projects of today has made this data warehousing technology simpler for use. Therefore, it is now much easier for software developers to master and implement in their new projects. The operations for data processing are simplified and their use is very simple in the majority of modern program languages. Moreover, it can be concluded that the data can be analysed and displayed to the users in real time using contemporary techniques and technologies in programming, which is a great advantage when the management needs to make critical decisions.

Likewise, this paper has presented warehousing of big files using GridFS. Today, when web 2.0 applications almost do not contain text data, managing big files is crucially important. GridFS represents MongoDB's solution for storing binary files into a database. As it could be seen in this paper, the solution offers a number of benefits when working with big files, from storing files in a distributed architecture, safety, possibility of storing metadata, etc.

As it can be concluded from the facts mentioned in this paper, NoSQL databases are definitely a good option in data storing in business organizations nowadays. The multiple benefits they offer are the reason why companies increasingly implement these databases in their work environment. Furthermore, the emergence of new technologies like Internet of Things (IoT), which use sensor networks to collect numerous data from their environment, make SQL an ideal solution for data storing.

Storing and analysing of these data offers a new way of collecting and reviewing new information to business organizations, which can lead to new vision in their operation.

6. References

1. B. W. Wirtz, O. Schilke and S. Ullrich, „Strategic Development of Business Models: Implications of the Web 2.0 for Creating Value on the Internet“ *ScienceDirect Long Range Planning*, vol. 43, Issues 2–3, pp. 272–290, April–June 2010.
2. Cisco Co., „Cisco Visual Networking Index: Forecast and Methodology, 2009–2014“, *Cisco Public Information*, June 2010.
3. J. McKendrick, „Big Data Market Set to Explode This Year, but What Is 'Big Data'?“, SmartPlanet.com, February 2012.
4. M.P. Papazoglou, „Service-oriented computing: concepts, characteristics and directions“, *Web Information Systems Engineering, 2003. WISE 2003.*, pp. 3 – 12, December 2013.
5. Elragal, „ERP and Big Data: The Inept Couple“, *ScienceDirect on Procedia Technology CENTERIS 2014*, vol. 16, pp. 242–249, Novmer 2014.
6. J. Pokorný, „New database architectures: Steps towards big data processing“, *IADIS European Conference Data Mining 2013*, 2013.
7. DataStax, „Why NoSQL?“, October 2012.
8. B. M. Moniruzzaman and S. A. Hossain, „NoSQL Database: New Era of Databases for Big data Analytics – Classification, Characteristics and Comparison“, *International Journal of Database Theory and Application*, vol. 6, No. 4., 2013

9. R. Cattell., „Scalable SQL and NoSQL data stores“. *ACM SIGMOD*, vol. 39, Issue 4, pp 12-27., December 2010.
10. Bhatewara and K. Waghmare, „Improving Network Scalability Using NoSql Database“, *International Journal of Advanced Computer Research*, vol. 2, No. 4, Issue 6, pp. 488-490, December 2012.
11. Oussous, F. Benjelloun, A. A. Lahcen and S. Belfkih, „Comparison and Classification of NoSQL Databases for Big Data“, *International Journal of Database Theory and Application*, vol. 6, No. 4.2013, 2013.
12. J. Pokorny, „NoSQL databases: a step to database scalability in web environment“, Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services, pp. 278-283, December 2011.
13. BĂZĂR, „The Transition from RDBMS to NoSQL. A Comparative Analysis of Three Popular Non-Relational Solutions: Cassandra, MongoDB and Couchbase“, *Database Systems Journal*, vol. V, no. 2/2014, March 2014.
14. N. Leavitt, „Will NoSQL Databases Live Up to Their Promise?“, *IEEE Computer*, vol. 43 , Iss. 2, pp 12-14, Feb. 2010.
15. E. Lotfya, A. I. Salehb, H. A. El-Ghareebc and H. A. Alid, „A middle layer solution to support ACID properties for NoSQL databases“, *Journal of King Saud University - Computer and Information Sciences*, vol 28., Iss. 1., pp. 133–145, January 2016.
16. J. McKendrick, „Big Data Market Set to Explode This Year, but What Is ‘Big Data’?“, http://www.mckinsey.com, Februar 2012.
17. P. Hofmann, „ERP is Dead, Long Live ERP“, *IEEE Internet Computing*, vol. 12, Iss: 4, pp 84-88, July-Aug. 2008.
18. K. Chodorow, *MongoDB: The Definitive Guide*, O’Reilly Media, USA, pp 389-390, 2013.
19. Z. Parker, S. Poe and V. S. Vrbsky, „Comparing NoSQL MongoDB to an SQL DB“, *Proceedings of the 51st ACM Southeast Conference*, Article No. 5, April 2013.
20. Nayak, A. Poriya and D. Poojary, „Type of NOSQL Databases and its Comparison with Relational Databases“. *International Journal of Applied Information Systems (IJAIS)*, vol. 5, No.4, pp 16-19, March 2013.